Copyright © 1974 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 53, No. 6, July–August 1974
Printed in U.S.A.

Combining Intraframe and Frame-to-Frame Coding for Television

By J. O. LIMB, R. F. W. PEASE, and K. A. WALSH

(Manuscript received October 9, 1973)

A method of frame-to-frame coding is proposed in which the changes from one frame to the next are first detected and then transmitted as an intraframe coded signal rather than as frame-to-frame differences. A coder was constructed to test the proposal using DPCM for the intraframe encoding.

Three aspects of the coder design presented particular problems. They were:

- (*i*) *Movement detection (as a result of the increase in frame-to-frame noise caused by the intraframe coding).*
- (*ii*) *Smooth reduction of bit-rate and picture quality so as to take advantage of the reduction in spatial quality that a viewer tolerates when areas are moving fast.*
- (*iii*) *Control strategy for linking the operation of the buffer, the movement detector, and the operating state of the coder.*

The coder gave good picture quality at a transmission rate of 1.5 megabits per second (0.75 bit per picture element), except in extreme situations where the moving area covered almost the entire screen. The performance is described in detail at bit rates of 2.0, 1.5, and 0.5 megabits per second.

The experimental coder has a number of desirable properties from an overall systems point of view when compared with transmission of frame differences. These include high tolerance to transmission errors and small frame storage requirements.

I. INTRODUCTION

More than forty years ago it was first realized that channel capacity requirements could be significantly reduced by transmitting only those parts of a television signal that represent the changes from one frame of an image to the next.[1] However, only recently technology has been available to store a complete frame of video information to enable such a system to become practicable.[2,3]

In addition to the high correlation from frame to frame (temporal correlation), quite high correlation also exists from line to line and between adjacent elements along a line. It is these spatial forms of correlation which have been most widely exploited in coding television signals. For example, within a single frame we can switch between previous element prediction and previous line prediction, depending on whether there is more horizontal or more vertical similarity between adjacent picture elements.[4] Similarly, in frame-to-frame coding the element in the previous frame corresponding to the element being encoded is a good prediction when an object is moving slowly, whereas a spatially adjacent element in the same frame is a better prediction of the current element when the object is moving fast.

In an ideal situation, it is easy to determine the changeover point at which the element difference is smaller than the frame difference. Consider an image moving horizontally at a constant speed of one picture element per frame period (pef). This speed is quite slow; it would take about 8 seconds for an object to cross from one side of the screen to the other. During one frame an element moves so as to occupy the position occupied by the element adjacent to it in the previous frame. Consequently, at this speed the element-difference signal equals the frame-difference signal: at greater speeds the frame-difference signal is larger.[5]

One early scheme for frame-to-frame coding, called Conditional Picture-Element Replenishment, updated the changed picture elements with a new PCM value.[3] We refer to this as CR/PCM coding. The efficiency of this scheme can be improved significantly by transmitting the difference between a stored reference frame and the new frame (CR/FF). The changes can be transmitted with little more than four bits per element, on the average, rather than between six and eight bits for PCM transmission.[6]

In conditional replenishment (CR) schemes, data are generated at a very uneven rate, and therefore it becomes necessary to use a buffer to smooth the peaks if a constant transmission bit rate is required. In general, while the buffer can smooth data within the field, it is not practicable to smooth from one activity peak to the next because the size of the buffer would need to be very large.* Further, in the video-

---

* For example, if a movement lasted for a duration of 1 second, between 3 and 6 megabits of data could easily be generated, most of which would need to be stored (Ref. 7).

telephone situation, the signal delay inherent in a large buffer becomes intolerable to a user. Consequently, the efficiency of a coding scheme is highly dependent on the peak data generation rate. However, the coding of moving areas by intraframe techniques becomes more efficient with faster movement. This is in contrast to most other frame-to-frame coding schemes in which the efficiency decreases with the speed of movement. There are other advantages to coding the moving parts as an intraframe signal:

(*i*) In many video-telephone situations, only the intraframe coded signal is available and, in general, transmitting the intraframe signal minimizes requantization effects.

(*ii*) Such a scheme lends itself very well to economizing on frame storage requirements by storing only intraframe differences.

A conditional replenishment system using intraframe coding of the changed parts of the signal (CR/IR) was first demonstrated in 1970.[8] This paper describes that system and subsequent improvements associated with movement detection and the control strategy. Related work is described by Wendt[9] and Kanaya is currently investigating a CR/IR type system.[10]

The concept of CR/IR coding is illustrated in Fig. 1. The output of an intraframe coder is stored locally in a frame-memory loop. If a significant difference is detected between the input signal and the decoded version of the stored signal, the two switches move to position 1 and new data are entered into the frame memory and at the same time transmitted to a frame memory at the receiver. It is also necessary to

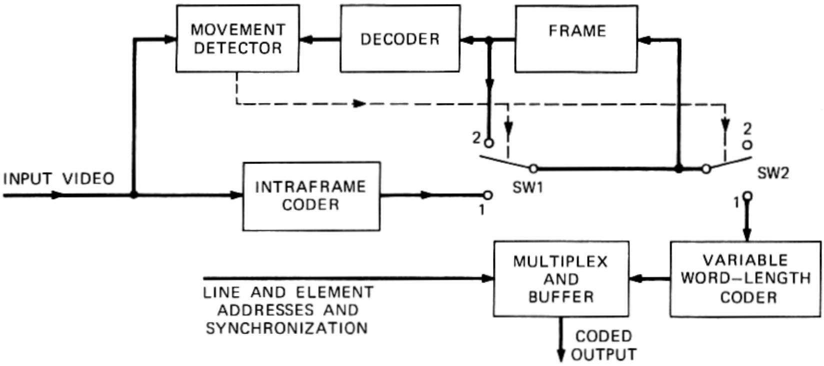


Fig. 1—Basic concept of the conditional replenishment intraframe (CR/IR) coder.

transmit addresses so that the receiver can insert the coded signal in the correct place.

Figure 1 is deceptively simple, and a large combination of techniques is needed to implement such a coder successfully. However, the configuration we describe should not be regarded as a complete system, but rather as the result of an experiment, first, to evaluate the feasibility of transmitting an intraframe-coded signal in moving areas and, second, to explore methods of varying and controlling the horizontal accuracy with which the intraframe signal is coded.

A brief description of the CR/IR coder is given in the next section, while more details are given in the appendix, Section A.1. Section III describes the performance of the coder and Section IV discusses, first, some additional techniques which could be used for further improvement and, second, some implications of CR/IR coding for overall system design.

II. DESCRIPTION OF CONDITIONAL REPLENISHMENT INTRAFRAME CODING TECHNIQUES

2.1 Switching between "stationary" and "moving" signals

Let us be specific and assume that the intraframe coder is a differential quantizer[11] (differential pulse-code-modulation coder). The scheme of Fig. 1 works satisfactorily if the switch is operated (closed or opened) only when the digital value of the decoded form of the coded signal is the same at both the output of the frame memory and the output of the intraframe coder. If this condition is not met, an error term is added to the coded signal which is equal to the difference between the decoded value of the two signals incident at switch 1 at the instant of switching. This would result in a streaky picture with streaks similar to those produced by transmission errors. Figure 2 illustrates this lack of tracking between the intraframe coder and the CR/IR decoder when the switches of Fig. 1 change position to accept new data.

To permit the switches to change position only when there is no difference (or a very small difference) between the decoded values of the two signals arriving at switch 1 (Fig. 1) would be very restrictive and would probably result in a significant increase in the area to be transmitted, particularly if the input signal is at all noisy.

This difficulty is overcome with the configuration of Fig. 3.* The switch now handles normal (accumulated PCM) values rather than

---

* Notice that the input to the coder is in intraframe coded form. We imagine the CR/IR coder as being one stage of a hierarchy of coders in which each stage would probably be at different physical locations (Section 4.2).

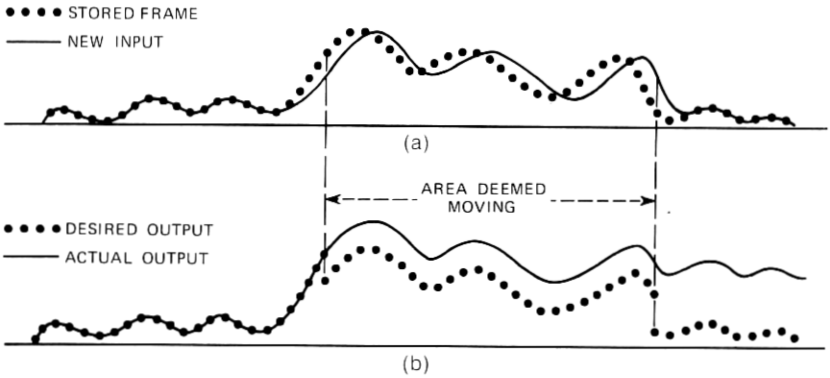


Fig. 2—Waveforms showing operation of conditional replenishment coder of Fig. 1. (a) Dotted line: Decoded value of stored signal (in frame memory); solid line: New incoming signal which is shifted to the right in the moving area because of a change in position of subject. (b) Solid line: Output of conditional coder of Fig. 1. Notice the offset at the instant of switching caused by addition of a new element-difference signal to the old (stored) decoded signal; dotted line: Desired representation of the combined input and stored signals.

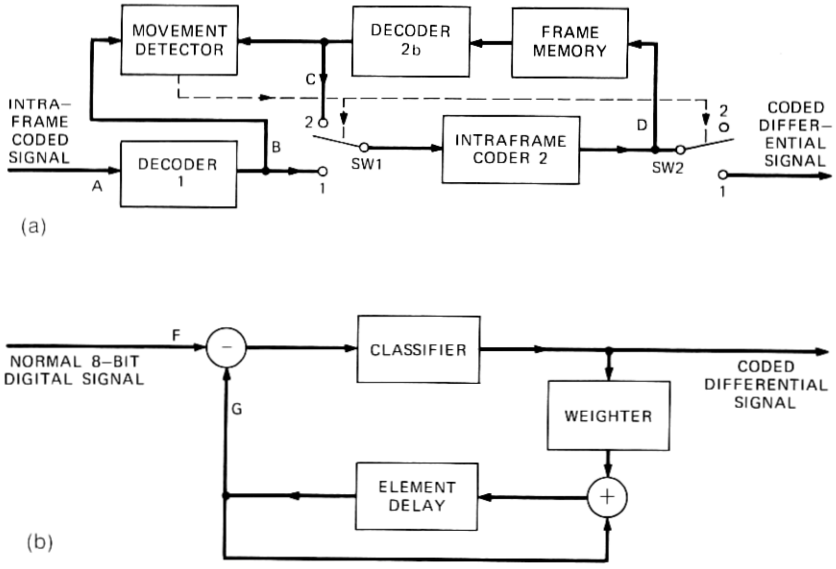


Fig. 3—(a) Diagram of the CR/IR coder. Notice the change from Fig. 1: Coder 2 and decoder 2b are added to the loop so that the offset problem shown in Fig. 2 is eliminated. (b) Diagram of intraframe coder (DPCM).

differential values, and the signal is recoded before it is stored in the frame memory. While there is no detected movement, the signal circulates through the frame memory, decoder, and coder without change. If the switch changes position while the PCM values entering switch 1 are identical, then the coded signal will not be changed after passing through decoder 1 and being recoded (i.e., the signals at A and D will be the same). On the other hand, if the switch changes when the two PCM values are different, a small amount of recoding noise will occur while the new signal is corrected.

The operation of Fig. 3 is probably best appreciated by a numerical example shown in Table I. Let us assume a five-level differential quantizer with decision levels ± 1, ± 4, and representative levels 0, ± 2, ± 6 (see, for example, Ref. 12).

If row A represents the input to decoder 1, then row B represents the output given that the value of the accumulator is 32 before decoding.* Let C represent decoder 2b output. Row D represents the output of intraframe coder 2 and is the same as the output from the frame memory before decoding up to the point that the switches change from position 2 to position 1. Row E is the accumulated value of D and represents the signal at the receiver. Just before switching, the difference between the two values B and C at switch 1 is 6. After switching, the input to intraframe coder 2 is 44 (signal F, Fig. 3(b)), while the value in the accumulator is 36 (signal G). The difference is +8 which is coded as a 6 (therefore, E is 42). Coding continues with $B = F$ as the input and row D as the coder output. On the fourth sample after switching, the two signals B and E are the same, and signals A and D will remain locked together until the switch returns to position 2. Thus, the coding noise at the switching point is confined to three samples and has the values −2, +2, +2 (obtained by subtracting row B from row E). The time to lock in depends on the quantizing characteristic, the input waveform, and the amount of difference at the instant of switching; in many instances, lock-in is immediate. The average lock-in time for the quantizer used in this study was measured

---

* We are dealing with many different types of signals in connection with the differential quantizer, and it is important to have a clear description of the terms used. A signal can be either analog or digital (i.e., PCM). The signal will be called "normal" (e.g., normal digital) if it is directly related to the amplitude of the video signal (signals B and C of Fig. 3(a)). Similarly, a signal will be called differential if it is directly related to some form of difference-signal (signals A and D of Fig. 3(a)). A standard 8-bit PCM signal will be referred to as a normal digital signal. A signal that has passed through both a differential coder and decoder will be called normal-differentially-quantized (signal B of Fig. 3(a)), while if it has only passed through the coder it will be referred to as a coded-differential signal (signals A and D of Fig. 3(a)).

Table I — Numerical example of the operation of the coder shown in Fig. 3

	Successive Picture Elements in Horizontal Direction										
A		+2	+2	0	+6	+2	+2	−6	0	0	+2
B	32	34	36	36	42	44	46	40	40	40	42
C	32	34	36	36	36	X	X	X	X	X	X
D		+2	+2	0	0	+6	+6	−6	−2	0	+2
E	32	34	36	36	36	42	48	42	40	40	42

↑
Switching Point

at 0.11 element per transition of switch 1 (Fig. 3(a)) for the case when there was virtually no movement and the small amount of updating that was occurring was triggered primarily by noise. Where there was a significant amount of movement, the average lock-in was 0.30 element.[*] A similar lock-in time is required when the switch 1 moves from 2 to 1 (return to stored signal).

2.2 Moving area detection

Accurate detection of changed areas within the picture is important for efficient coding. This is straightforward when working with a high-quality digital signal.[13,6] However, as can be seen from Fig. 3, we are detecting the changed areas from a signal that has been intraframe coded and is therefore relatively noisy, particularly at edges where the coarse outer levels of the differential quantizer are used. This means that more sophisticated movement-detection techniques are required to obtain adequate detection. Reference 14 derives some correlation properties of the types of frame-difference signals generated in conditional replenishment encoding and Ref. 15 describes the implementation of a previous design. The movement detection used in this study is similar in principle to that described in Ref. 15. The difference between the stored frame and the current frame is: (*i*) spatially and temporally filtered; (*ii*) applied to a varying threshold which is under control of a modified element-difference signal (this compensates for the larger errors introduced in high-detail areas by the differential quantizer) and (*iii*) "blocked-in," an operation which both produces a more contiguous moving area and rejects small isolated changes.

---

[*] These figures were obtained with the coder control circuit locked in mode 1 for the first figure ("no movement") and mode 2 for the second figure ("movement"). See Section 2.4 for a description of the various operating modes. We suspect that the short lock-in times result partly from the fact that the second representative level is twice the value of the first (see Table IV).

Specific details of the movement-detector used for this study are given in appendix Section A.2.

2.3 Reduction of resolution

As the speed of a moving object increases, the resolution of the resulting image in the direction of movement decreases because of the light-integrating action of the camera target. For horizontal movement, this in turn reduces the amplitude of the element-to-element differences, and the entropy of the associated intraframe coded signal decreases. Figure 4 is a picture of the unquantized element-difference signal of a moving object against a stationary background at two different speeds. The reduction in contrast is quite obvious in the moving area as the speed goes from one-half element per frame to four elements per frame.

Although it appears that the eye can detect smearing of the picture because of camera target integration, an observer is reasonably tolerant of this type of degradation and, in fact, we would like to take the process a little further. As we can see from Fig. 4 (see also Ref. 14), the effect of target integration is to reduce the bandwidth of the spatial signal in moving areas; this, in turn, reduces the first-order entropy of the coded signal. But relying solely on the first-order entropy reduction of the intraframe coded signal at full sampling rate does not take full advantage of the redundancy in the signal at high speeds, when the signal is essentially oversampled.*

Smoothly reducing the sampling rate as the speed increases would be very effective but is impracticable. Switching to a submultiple of the sampling rate is quite practicable, but the difference in picture quality in going from the full sampling rate to half sampling rate is quite large, especially for differential quantization. Thus, the change in quality at the instant of switching is noticeable.

A coding technique called receiver-model coding was developed partially for this application.[17] It enables properties of the observer to be incorporated into the coding process. A particularly simple form of receiver-model coding (referred to as "level variable sampling" in Ref. 18) is 2:1 horizontal conditional subsampling, in which every second point in the picture is differentially quantized in the normal manner. The alternate points (conditional points) are extrapolated from the previous point (zero-order hold) unless the error incurred by so doing exceeds a predetermined threshold (in which case, they are

---

* The results of Bobilin show how rise-time and edge-busyness change as the ratio between sample rate and bits per sample (for DPCM) is altered (Ref. 16).

(a)

(b)

Fig. 4—Reduction in amplitude of element differences with increase in speed. (a) Head moving at a speed of 0.5 pefs. (b) Head moving at a speed of 4.0 pefs. Reduction is caused by integration of light falling on camera target for duration of one frame.

also differentially quantized in the normal manner). When the threshold is low, nearly all points are coded normally. As the threshold is increased, more and more conditional points are extrapolated until, if the threshold is high enough, the signal is effectively subsampled. To have a bit-rate advantage with horizontal conditional subsampling, we need to use a variable-length code since information is transmitted about all points, including the conditional points unless the signal is fully subsampled (see appendix Section A.3.1).

The coder used in this study did not have a continuous threshold control, but could be switched to give one of five "operating states" starting with normal differential quantization and going to 4:1 horizontal subsampling, which gave a picture quality that was scarcely adequate even in very fast moving areas.

2.4 Control strategy

There are two different ways in which the data-generation rate may be reduced. One is by reducing the accuracy and resolution with which the moving area is coded as described above. The other method is to reduce the size of the moving area by demanding that the difference (measured in some way) between the stored signal and the incoming signal in a given area be larger before that area is regarded as moving. Raising the criteria for movement detection is most effective for areas that are moving slowly.

Two possible control strategies are;

(*i*) Use a measure of the speed of the moving object in the picture to reduce the resolution and, therefore, the data generation rate in the moving areas, but not so much that picture quality will be significantly affected. Data may still be generated at a rate that exceeds the channel rate, especially when large areas are moving slowly.

(*ii*) Use a measure of the buffer fullness to reduce the resolution and size of the moving area.*

At the time of this study, a speed-measurement circuit was not available and so the buffer alone was used to control both the spatial resolution within the moving area and the size of the moving area.†

---

* These types of control are quite different in effect (Ref. 7).

† Some relatively simple techniques for determining the approximate speed of the moving area are currently being evaluated by the first author and J. A. Murphy of Bell Laboratories.

Table II — Bit-rate control modes—summary of the bit-rate reduction techniques for each mode

Mode (Section A.3.4)	Level Variable Sampling (Section A.3.1)				Moving Area Detector Threshold Select (Section A.3.2)					Single Point Threshold (Section A.3.2)	
	Levels deleted 2:1			4:1							
	±1	±1 and ±2	All	All	T_1	T_2	T_3	T_4	T_5	Low	High
1					X					X	
2	X					X				X	
3	X						X			X	
4		X					X			X	
5			X				X				X
6			X					X			X
7				X					X		X
8	Frame Repeat										

Feedback from the buffer progressively reduces spatial resolution and increases thresholds for moving area detection in a sequence of eight steps with the last step being the prevention of all updating.

We have built a system based on the scheme of Fig. 3 using a simulated buffer with the buffer-control strategy described above. The equipment is described in detail in the appendix, and the feedback modes are summarized in Table II. The experiments carried out and the results obtained are described below.

III. EXPERIMENTS AND RESULTS

The functional blocks of the coder interact in a complex manner, making it difficult to evaluate the separate contribution of each block. Furthermore, transitions between modes can occur very rapidly so that in certain instances the coder may oscillate between adjacent modes at line rate. We first report the performance (picture quality and bit rate) of the operating states applied to the whole picture (with no movement detection or feedback control). Next, we describe the additional effect of movement detection still without feedback control. Finally, we describe the performance of the overall coder at different transmission rates.

A head-and-shoulders view was used with the subject covering slightly less than half of the viewing area. Thus, with the size of the subject constant, varying the speed at which he or she moved across the screen varied the data rate. The subject was wearing relatively low-detail clothing; when high-detail clothing is worn, the data rates are a little higher.

3.1 Resolution reduction: effect of changing operating states

Table IIIa shows the performance of the coder with the various operating states applied to the whole of a stationary picture. The bit rate represents the amplitude bits per picture element and, of course, does not include addressing, etc. There is a bit-rate reduction of 45 percent in going from full sampling to 2:1 sampling accompanied by a gradual decrease in picture quality.

3.2 Effect of moving area detector

To show the effect on bit rate of each mode (described in Table II), the speed of a subject was chosen so that when only mode 1 is used (feedback-control inhibited and manually selecting mode 1), the bit rate needed for transmission was approximately 2.0 megabits per second. While the subject conditions are kept constant, each remaining mode was manually activated and the resulting bit rate recorded (Table IIIb). Here the bit rate is a total system bit rate (appendix Section A.3.3). There is about a 10:1 drop in average bit rate in going from mode 1 to mode 7. The reduction in bit rate in going from mode 2 to mode 3 and from mode 5 to mode 6 is a result only of a reduction in the moving area (see Table II). These measurements are not an exact indication of the bit-rate reduction of each mode, since in actual

Table IIIa — Bit rate and picture quality for each operating state with movement detection disconnected (coding applied to whole picture)

Operating State	Level Deletion	Bit Rate (bits/picture element)	Picture Quality
1	None	3.07	Very good. Limited only by the quantization process.
2	Level ± 1	2.58	Very good. There is a just-noticeable increase in noise in areas having fine detail and low contrast.
3	Levels ± 1 and ± 2	2.36	Good. The increase in random noise is more noticeable than state 2, and some fine detail with low contrast is lost.
4	2:1 subsampling	1.69	Fair. Sharp edges become serrated and fine detail is blurred.
5	4:1 subsampling	0.89	Poor.

Table IIIb — Bit rate for each control mode

Mode	1	2	3	4	5	6	7
Bit rate (Mbits/s)	2.01	1.60	0.88	0.80	0.64	0.46	0.19

operation the speed and size of the moving area would be different for each mode.

3.3 Performance at different transmission rates

3.3.1 Performance at 1.5 megabits per second

Table IIIc gives the performance of the system operating at a transmission rate of 1.5 megabits per second. To enable detailed observation and measurement of the effect of each mode, the coder was locked to each mode. Then the picture quality and amplitude bits per transmitted element were recorded for the type of movement appropriate to that mode. The picture quality depends strongly on the size of the moving area; as noted, the moving subject filled approximately half the picture. With smaller moving areas, the higher modes are used less frequently and the picture quality is better; the situation reverses in larger moving areas. In the table, conversational movements are considered movements of the face and gentle head movements. The X denotes that these modes cannot be activated only by side-to-side body motion.

3.3.2 Performance at 2.0 megabits per second and 500 kilobits per second

With the coder operating normally, the picture quality was observed at transmission rates of 2.0 megabits per second and 500 kilobits per second.

At 2.0 megabits per second, very slow (1 pef) to moderate (3 pef) side-to-side movements cause mode 1 to be used continuously. This provides good picture quality and also good moving area detection. Only during very fast motion does mode 3 come into use, which reduces the accuracy of the moving area detection and subsampling on the inner pair of levels. Mode 5 is used only for violent changes such as panning the camera or walking in front of the camera. The noticeable defect is a coarse structured effect in the moving areas produced by the 2:1 subsampling and the reduced accuracy of the moving area detector.

Table IIIc — Picture quality and bit rate for each control mode at 1.5 megabits per second

Mode	Amplitude Bits per Transmitted pel	Mode Activated by		Buffer Level (fraction of full capacity)	Picture Quality
		Speed of Movement (pef)	Type of Movement		
1	3.3	$<\frac{1}{2}$	Normal conversational	$<\frac{1}{8}$	Very good. Slight increase in edge noise at very slow speeds because of movement detector. Picture quality is better than with normal DPCM because the quantizing noise is less visible when it is "frozen."
2	2.9	$1\frac{1}{2}$	Active conversational	$\frac{1}{8}$	Good. A higher moving area detector threshold is chosen, and thus the area detected is reduced. The result is a slight increase in low detail noise.
3	—	3	Active conversational with hand and arm movements	$\frac{1}{4}$	Fair. A higher threshold and the elimination of temporal feedback cause some low detail areas in motion to appear somewhat contoured or "patchy." These effects are only marginally noticeable.
4	2.2	5	Very active body, hand, and arm movements	$\frac{3}{8}$	Fair. The threshold is the same as above; however, the added noise caused by the level variable sampling causes movement detection to become slightly worse.
5	1.7	X	Violent motion or standing up in front of camera	$\frac{1}{2}$	Acceptable. The threshold remains the same, but the added noise of the sampling and the FOS inhibit (see section A.3.1) reduce movement-detection accuracy. Also, at some speeds (multiples of 2 pef) the sampling produces a striped structure.
6	—	X	Motion such as walking in front of camera	$\frac{5}{8}$	Marginally acceptable. A higher threshold is chosen which produces a noticeable dirty window effect.
7	—	X	Motion such as panning the camera	$\frac{7}{8}$	Poor.

With the transmission rate limited to 500 kilobits per second and the subject in very slow side-to-side motion (0.5 pef) or in normal conversational movements (i.e., gentle lip and head movements), the system uses only the first four modes and the quality picture is still good. At a speed of 1 pef, modes 5 and 6 are used in which 2:1 subsampling is employed and the movement-detector uses the higher thresholds. The result is a slightly more noisy picture with the movement detector producing either a "dirty window" or a patchy effect.

At a speed of 2 pefs, mode 6 is mostly used. At this point the picture quality is probably unacceptable with the major degradations being: (*i*) the coarse structured effect caused by poor movement detection, (*ii*) the noisy edges caused by the 2:1 subsampling, and (*iii*) the general increase in noise.

At a speed of 3 pef, mode 7 is used more frequently and the picture becomes unacceptable, with the major degradations being poor moving area detection and a "column" effect produced at some speeds by the 4:1 subsampling.

IV. DISCUSSION

The above experiments are only a start in investigating the techniques of CR/IR coding. However, even at this stage we can see the encouraging performance for fast moving scenes. For example, at a transmission rate of 2 megabits per second, motion such as panning the camera only invokes mode 5; i.e., neither 4:1 subsampling nor the highest levels of the movement detector are used. In a previously described CR/FF coder, motion such as panning the camera invoked frame repeating.[6] Further work is needed to examine related techniques that could significantly improve coder performance. One example is an evaluation of intraframe coding techniques that are more efficient and better suited to CR/IR operation. In addition, we should investigate the application of known frame-to-frame coding techniques; we discuss some of these below.

4.1 Add-on techniques

The vertical resolution can be reduced by transmitting only alternate lines in each field and filling in the missing lines by vertically averaging. In this study, the horizontal resolution was reduced by up to a factor of 4. This is inferior to spreading the resolution reduction more equally between the vertical and horizontal dimensions. A horizontal resolution reduction of 4:1 is acceptable in very fast moving areas, but if the mode is invoked at lower speeds, for example, where the camera is

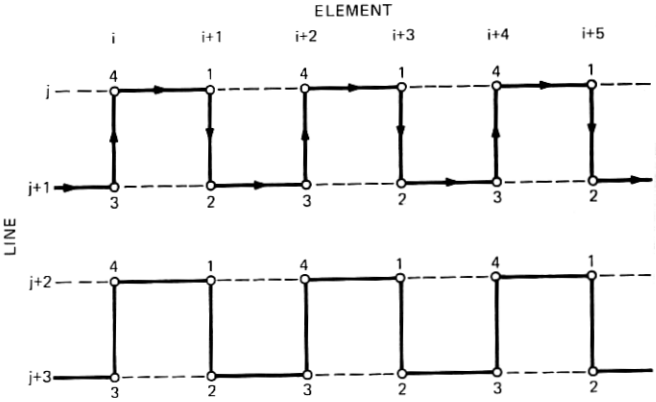


Fig. 5—Merli scan path in which elements are taken alternately from adjacent scan lines. Picture elements are processed in quads.

being panned slowly, then serious degradation results. One method for smoothly reducing the resolution in both dimensions by a combined factor of 4 would be to apply receiver-model coding to the Merli scanning algorithm.[17,19] In this coding scheme two adjacent scan lines are coded simultaneously by following the notched path of Fig. 5. The elements are processed in quads with the number 1 elements always coded with full precision. An attempt is made to represent the number 2, 3, and 4 elements as linear interpolations based only on the number 1 elements. The interpolation error is calculated and filtered to approximate the liminal vision of the human observer. If the filtered error signal at a particular point exceeds the allowed threshold for a given quality, then the point is updated.

As the threshold is raised, fewer conditional elements (numbers 2, 3, and 4) are transmitted. If the threshold is raised far enough, a 4:1 subsampled picture is obtained with a reduction of 2:1 in both the vertical and horizontal directions. Horizontal subsampling reduces the number of amplitude bits that have to be transmitted without affecting the number of address bits or line synchronizing bits.* By using the Merli algorithm, on the other hand, the line address and synchronizing bits would be almost halved since a line now contains twice as many elements as it previously did.

Conditional-vertical subsampling is a technique that is applied from field to field. Alternate fields are obtained by a four-way average of

---

* Actually, one bit could be dropped from the address word when 2:1 subsampling is used.

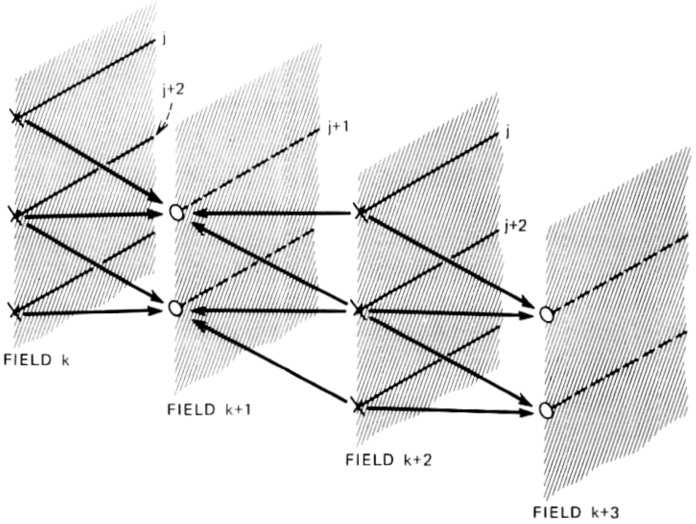


Fig. 6—Four-way averaging in which alternate fields are not transmitted. At the receiver, the missing fields (even-numbered fields) are replaced by a four-way average of elements in the adjacent fields.

samples in the immediately preceding and succeeding fields as shown in Fig. 6. Should the average fail badly for a particular element, then an additional correction signal may be transmitted, depending on the quality that is required. This four-way field averaging reduces both spatial and temporal resolution by a small amount.[20]

More severe temporal averaging can be employed by using what may be called conditional frame-to-frame subsampling. Such techniques are most useful where large areas are moving slowly, the particular condition which is handled poorly by the CR/IR coder and quite easily by the CR/FF encoder. However, if there is a significant reduction in temporal resolution, it is important that it be under the control of a speed-indicator circuit so that it can be switched out when the speed starts to increase.

4.2 System implications

In a practical visual communication system, transmission links will vary greatly in length. As a consequence, on short links a simple inexpensive coder would be appropriate, whereas on longer links more expensive frame-to-frame encoding might be suitable. Now in a complex switched system, we may well want to pass through a number of digital links in tandem, some being short and others long. Thus, it is important to have a family of coders that are compatible in the sense that they can operate in tandem without unduly degrading the system. We could envisage at least four stages of coding: (*i*) a simple differential quantizer stage; (*ii*) a more efficient intraframe encoder using a variable-length code on the output of 1; (*iii*) an interframe coding stage and (*iv*) a channel-sharing stage where a number of users share a high capacity channel, trading on the fact that there is a low probability of all users being active simultaneously (as in TASI).[21,22] The conditional intraframe coder is well suited for this type of multistage tandem operation. As we have seen, the frame-to-frame coding stage does not add quantizing noise to the signal except in elements adjacent to the points of switching between stationary and moving areas or when feedback from the buffer decreases the accuracy of the intraframe coder in the storage loop. If the signal is converted back to the intraframe form and frame-to-frame encoded for a second time, then the second frame-to-frame encoding will give a signal that is identical to the first frame-to-frame encoding if one prerequisite is met: the position of the switching points between moving and stationary areas are indicated in the intraframe signal. This would increase the intraframe data rate by approximately 2 percent.

If an improvement is made in the performance of the intraframe encoding stage, this improvement will carry right through to the frame-to-frame channel-sharing stages.*

The fact that an intraframe coder is connected to a CR/IR coder will tend to affect the type of algorithms that we employ in the intraframe stage. For example, techniques that complicate the encoder design but require a simple decoder will be preferred because there are more decoders in the system than there are encoders (see Fig. 3). Notice that the conditional horizontal subsampling studied here requires no modification of the decoder design.

4.2.1 Feedback control

The different coding stages of the overall coding hierarchy would normally be at different switching offices. This almost certainly rules out any feedback from one stage to a previous stage of coding, since to incorporate feedback would considerably increase the overall complexity. For this reason, the feedback control to achieve level-deletion was kept within the frame-to-frame coder (Coder 2 of Fig. 3) rather

---

* Of course, changes in the intraframe encoder may well necessitate changes in the encode and decode blocks of the frame-to-frame coder.

than operating on the primary encoder. There are two consequences of this restriction for the simple type of receiver-model coding employed here. First, the effective threshold used to delete components must jump from decision-level to decision-level rather than increase smoothly because by precoding the signal in the primary encoder the element-to-element changes are restricted to the small set of values allowed by the differential quantizer. Second, there is a small increase in coding noise since the two tandem intraframe encodings are different when level-deletion is used in coder 2. In practice, however, the smoothness of control is quite adequate.* The increase in coding noise when compared with feedback to the primary encoding stage is just noticeable in a stationary picture but is virtually impossible to detect in the operation of the overall system.

Recoding noise resulting from feedback control could become a problem with, for example, higher quality systems. However, there are intraframe coders that would virtually eliminate the problem. These coders transmit two or more separate signals which represent different components of the signal so that when one component is deleted the coding of the other component is unaffected. In one system of this type,[23,24] every second sample is transmitted as PCM or DPCM and the alternate samples are transmitted as a correction signal between an estimate based on the first set of signals and the actual input. Thus, the correction signal may be deleted without interfering with the coding of the main signal. Another example of such an encoding is the Hadamard transformation applied to a small block of picture elements;[25] higher-order components can be deleted without interfering with the decoding of the lower-order components.

4.2.2 Error performance

In achieving the improved performance of CR/FF coding over CR/PCM coding, certain system advantages were lost. These advantages are partially regained with CR/IR coding. Consider, first, the effect of transmission errors on picture quality.

Since a separate interframe decoder has not been constructed, experiments on the behavior of the CR/IR coder-decoder in the presence of channel errors have not been possible. However, some intuitive predictions can be made by considering the effect of different types of errors.

---

* We only used two intermediate steps (level ±1 delete, level ±1 and ±2 delete) out of a possible six.

If an amplitude word (as distinct from an address word) is in error, a noise streak will be introduced into the picture which will probably extend to the end of a line unless predictor leak is used. When there is a lot of movement there is a high probability that the error will be eliminated in the next frame since, by the usual nature of movement, the segment in error will likely be updated in the next frame and the updated segment builds only on information in the corresponding line of the stored frame to the left of the segment. With no movement or slow movement, there is much less chance that a segment in error will be "written over" in the next frame and the line in error would persist in the picture.

The signal can be made significantly more robust by transmitting a six- or seven-bit normal digital signal value at the start of a segment along with the addressing. In this way, updated segments would not build on the past values in any way. Based on an average of three segments per line, the additional amplitudes would require 0.145 megabit per second. The transmission of the additional values would terminate the effect of transmission errors already introduced and, by comparing the amplitude with the decoded value, errors could be detected. Once detected, substitution techniques could replace the line in error with a best estimate. This estimate would then last until the area was again updated. If, instead, the moving area addressing information is in error, then a large unpredictable section of a line will be in error. The effect of an error in the element address will be similar to an amplitude error, but on the average should affect a larger section of line.

In a practical system, we would want to send the line address word very securely and the start-of-frame word even more securely. The latter poses no problem since, as it occurs so rarely, it requires a negligible increase in bit rate to assign a large number of bits to the word.

It is interesting to consider what would happen if both frame and line synchronization were completely lost. Assume the receiver was aware of the loss and that it reset the frame memory to zero. Then, as soon as the person moved at the transmitting end the area in movement would be relayed faithfully to the receiver and the background would be inserted in the newly revealed area.

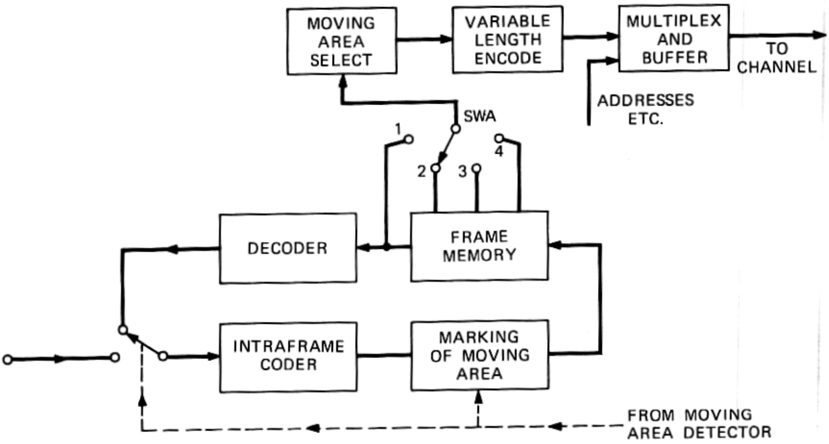
Although no experiments have yet been performed to determine channel error response, it appears that by transmitting an amplitude word before the start of each moving-area segment and using error detection and substitution techniques, the conditional intraframe encoder could be made to give acceptable performance at error rates as high as 10 or 20 per frame (an error rate of 2 to 4×10^{-4}). Forced updating would probably not be necessary.

4.2.3 Data interleaving

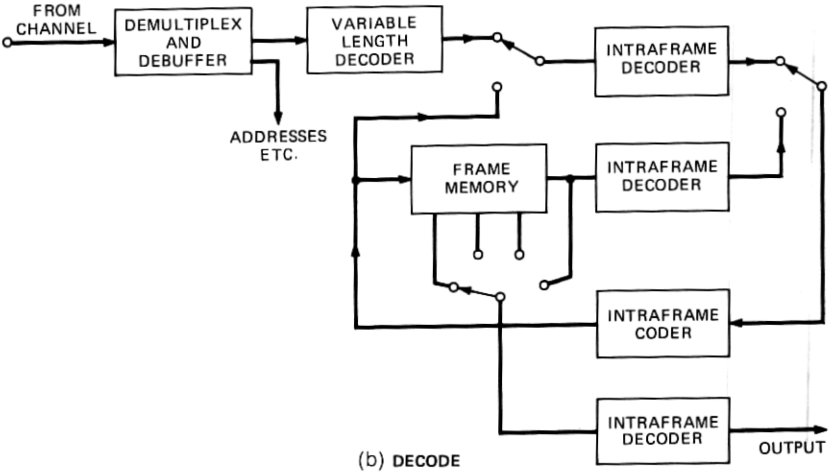
Data interleaving is a scheme for using the frame memory to achieve a degree of smoothing of the coded data, thus considerably reducing the size of the buffer store required or eliminating it altogether.[26] It has been shown that, unless a very large buffer is used, the main smoothing effect is already achieved with a buffer large enough to smooth the irregular data over a field.[7] A 4:1 interleaving of data is achieved, for example, by transmitting lines in the order 1, 65, 33, 97; 2, 66, 34, 98; 3, 67 ···. Now if the signal stored in the frame memory can easily be converted to the coded transmission signal, then taps can be placed on the frame memory and the data can be transmitted in an interleaved manner. Two examples of coders in which the signal is stored in the frame memory in a form similar to the transmitted signal is the CR/PCM coder and the CR/IR coder. Note, however, that the frame memory stores the whole picture and we need to know which components are to be transmitted. This information would have to be included in the stored signal and would probably result in a 5-percent increase in the size of the frame memory. If a four-bit word were used to represent the differential signal, one combination could be reserved to denote a change, either from a nonupdated to an updated segment or in the reverse direction.

Data interleaving is shown applied to the CR/IR coder and decoder in Fig. 7. Code words are inserted at the coder to denote changes between updated and nonupdated segments before the signal is stored in the frame memory; these words are disregarded by the local decoder. Switch A selects lines according to the required sequence and the moving area selector interprets the marker words and selects those segments for transmission that have been newly updated. The main decoder loop [Fig. 7(b)] operates on the data just as it is received; that is, the signal in the frame memory is stored in interleaved form. The signal is de-interleaved and decoded in order to obtain an output.

Notice that such a scheme would not work if the intraframe algorithm operated on more than one line at a time since the decoder is not processing consecutive lines. Such a restriction would not apply to the vertical processing of the Merli algorithm where a line, in essence, is twice as long as a normal line.



(a) ENCODE



(b) DECODE

Fig. 7—Data interleaving applied to the CR/IR coder. Data interleaving reduces buffer size by shifting much of the smoothing operation from buffer to frame memory. Note that data in the decoder frame memory are in interleaved form.

4.2.4 Channel sharing

Haskell has simulated a channel-sharing and buffering scheme in which a number of encoder outputs are combined and transmitted over one high data-rate channel with one large buffer.[22] He shows that in this way the channel requirements are more than halved when 20 encoder outputs are combined. In an actual system, in the unlikely event that a large number of users were simultaneously active, there would be feedback from the channel-sharing circuit to the encoders to reduce the data-generation rate by reducing picture quality in some manner. Although this would occur very rarely, the situation must be accommodated since we cannot arbitrarily discard data without seriously affecting picture quality: to ensure that the situation never occurs could require a significant increase in channel rate.*

Ideally, we would like to insert channel-sharing at multiple points in the transmission path, and these points may be quite remote from the encoder.[22] In this situation, feedback from the channel-sharing stage to the frame-to-frame coder would considerably complicate the overall system design. However, the CR/IR coder would enable data to be discarded with little effect on picture quality, since each new segment does not build on the past coded signal (assuming that a starting amplitude is transmitted with each segment as discussed in Section 4.2.2).

Thus, if overload of the channel-sharing stage were imminent, the whole line could be deleted except for the line addressing word (required for receiver synchronization) and a further special code word that would be inserted to inform the receiver that the line had been deleted. The receiver would then make a best estimate of the missing line based on the signal that it already has and the current control mode of the receiver (see, for example, Ref. 27). The line would be corrected by normal updating of the moving area. One would like to use a channel-sharing strategy that fairly evenly distributes deleted lines among the updated lines of all users and thus minimizes the possibility of deleting consecutive lines from one source.

4.3 Comments on conditional element-difference vs. conditional frame-difference coding

As mentioned in the introduction, transmission of element differences and transmission of frame differences are complementary in many ways. When transmitting frame differences, it is easier to control smoothly the temporal resolution since we are working directly with frame differences. We can still achieve a similar result when trans-

---

* The results of Haskell indicate that the variation in channel-rate requirements is only about 10 percent for 20 sources. However, there are a number of reasons why the variation could increase significantly in an actual system: (*i*) Channel-sharing schemes which minimize the buffering requirements would increase the variation. (*ii*) The interframe coders feeding the channel-sharing unit may be of different types. (*iii*) The channel-sharing unit may be designed for fewer sources or may have priority channels with different types of signal (e.g., data).

mitting element-difference signals: essentially the same signals as for frame-difference transmission are available to the transmitter on which to base a decision as to how elements should be coded. Similarly, it is easier to smoothly control spatial resolution when transmitting element differences, although we can achieve similar ends with frame-difference transmission (e.g., the horizontal subsampling used by Candy et al.).[6] Probably of most importance is the effect on the overall system of using one type of signal or another.

One tempting technique would be transmission of a frame difference in stationary or slowly moving areas and an element difference in fast moving areas or transmission of an element-difference-of-a-frame-difference.[9] Either method would tend to increase complexity associated with system considerations such as recoding, error mitigation, and channel sharing. It is also interesting that Wendt's results suggest to him that transmission of an intraframe coded signal is preferable to either transmission of a frame-to-frame coded signal or transmission of both signal types.

V. SUMMARY

We have described techniques for frame-to-frame coding in which the moving areas are transmitted as an *intraframe* coded signal (rather than as a PCM or frame-to-frame difference signal). This approach permits the intraframe encoding to efficiently adapt to the spatial resolution requirements of the moving area as the speed of an object changes. A coder has been constructed which uses a differential quantizer (DPCM coder) as the intraframe coder, and a strategy was developed for merging the new differentially quantized signal from the moving area with the old differentially quantized and stored signal from the stationary area with only transient error.

Because of inherent noise in the input signal and the error introduced in the initial coding, adequate detection of moving areas requires relatively complex processing involving a nonlinear, time-varying filter with an impulse response that extends temporally and spatially. The bit rate is kept within the capacity of the channel by feedback from the buffer to both the intraframe coder and the movement detection logic. As the buffer fills, the feedback reduces the accuracy of the intraframe encoding (and hence the bit rate) in four steps by a method referred to in an earlier paper as "level-variable sampling."[18] The feedback to the movement detector involves changing not only the level of significant frame-to-frame difference but also the parameters of the spatio-temporal filter contained in the movement detector.

The experimental study used a head-and-shoulders view occupying slightly less than half the field of view and a visible raster size of 255 lines by 220 elements. For a bit rate of 1.5 megabits per second, the picture quality sank below "fair" only for motion covering the entire field such as occurs when the subject stands up in front of the camera.

Transmission of an intraframe coded signal in the moving area leads to a number of advantages from the overall systems point of view when compared with the transmission of frame differences. By starting each transmitted segment within a line with a PCM value, updating becomes independent of previously transmitted data. Thus, errors will not propagate from frame to frame within the moving area. This also has implications for sharing a high-rate channel with a number of users where it would occasionally be necessary to delete segments of data. The signal is stored in the frame memory at the coder and decoder in intraframe coded form. This means that the frame memory need be only approximately half of that required to store the PCM signal. Further, since the stored signal can be simply converted to the form of the transmitted signal, we can use the data-interleaving technique to significantly reduce buffer requirements.

VI. ACKNOWLEDGMENTS

This study has in many ways grown out of the experience of, and discussions with, our colleagues over quite a long period of time. We express our thanks in particular to J. C. Candy, D. J. Connor, B. G. Haskell, F. W. Mounts, and W. G. Scholes.

APPENDIX

Description of Conditional Replenishment Intraframe (CR/IR) Coder

The picture format used in this study is similar to that used in the *Picturephone*® visual telephone system. There are 271 lines per frame, of which 255 are visible; 248 elements per line, of which 223 are visible; and 30 frames per second with 2:1 interlace.

The coding system that has been simulated consists of two parts, the primary intraframe encoding stage which is an element-differential quantizer and the secondary encoding stage which uses interframe techniques (Fig. 8). The output signal from the primary encoding stage is in normal differentially quantized form rather than coded differential form, thus avoiding the need for an additional decoder before the secondary encoder.

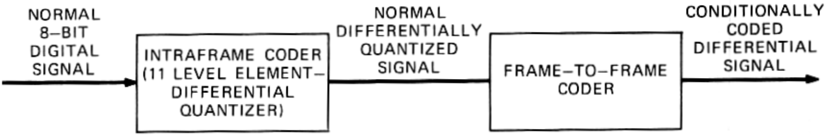


Fig. 8—Configuration of experimental CR/IR coder system.

A.1 Primary intraframe coder

The encoder used in this loop is an 11-level element-differential quantizer whose input, for our purposes, is a normal eight-bit digital (PCM) signal from an A/D converter, but in other respects is similar to that described in Ref. 12. In an actual system, the input would be analog rather than digital, but for experimental purposes it is more convenient to work with the digital signal. As shown in Fig. 9, it contains a decoder section whose output is the normal differentially quantized signal.

In the experiments to be described here, the accumulator loop has no "leak." However, the integrator is reset to a fixed value at the beginning of each line. The quantizer decision and representative levels are given in Table IV.

A.2 Movement detection

Since the outputs of decoder 1 and decoder 2b (Fig. 3) are separated in time by exactly one frame, they are used to form a frame-difference signal.* Frame differences caused by noise (negatively correlated in the moving area)[14] can be separated from those caused by motion

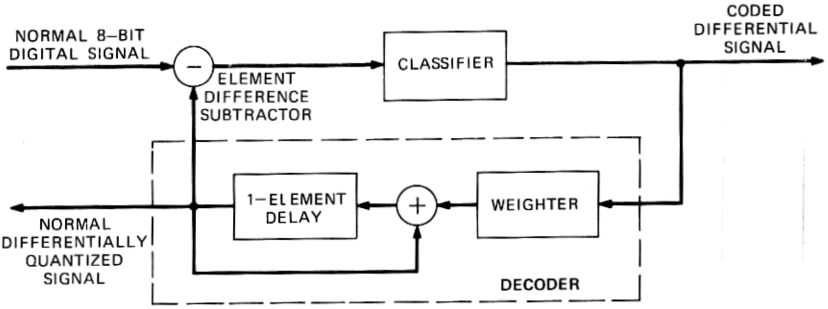


Fig. 9—Differential quantizer (DPCM coder) used as primary intraframe coder.

---

* For convenience, the term "frame difference signal" will be used, although it is actually the difference between a stored frame and a new frame, both of which have been differentially quantized.

Table IV — Quantizer level settings used by differential quantizer

Level No.	Decision Level (out of 256)	Representative Weight (out of 256)
0		0
	2	
± 1 inner levels		4
	6	
± 2		8
	12	
± 3		16
	22	
± 4		28
	36	
± 5 outer levels		44

(mostly positively correlated) by employing various spatial and temporal filtering operations. In addition, certain compensations are made for the nonlinear nature of the coding noise.

A block diagram of the moving area detector used in this experiment is shown in Fig. 10. The frame-difference signal is first fed to a spatial filter which provides a simple average over four adjacent elements along a line (4 × 1 filter). Temporal, single-pole filtering is then provided by placing the spatial filter in a feedback loop with a field delay. Since noise in the frame-difference signal is negatively correlated only in the updated area,[14] we would like to use temporal low-pass filtering only when updating occurs. This is achieved by closing the feedback loop (via switch 1) only when movement is detected. Since it would be expensive to delay a six-bit signal for the duration of one field, a different method was used. A three-bit dither signal was added (adder 1) to the output of the 4 × 1 spatial filter and the resulting sign-bit was used as a one-bit representation of the signal.* The field-delayed signals from the line above and below the current line are added and then assigned a "value" or "weight" before being added (adder 3) back into the frame-difference signal. The loop-gain, or the amount of temporal filtering, is controlled by means of the weighter. The spatio-temporal impulse response of this filter is rather unusual, spreading vertically as well as temporally and horizontally because a field delay, rather than a frame delay, is used (see Fig. 11).

The output of the spatio-temporal filter is then converted from 2's complement to sign-magnitude form (Fig. 10). A modified version of

---

* Other more complex one-bit representations could have been used; one-bit companded delta modulation, four-bit PCM samples at one-fourth of the sample rate.

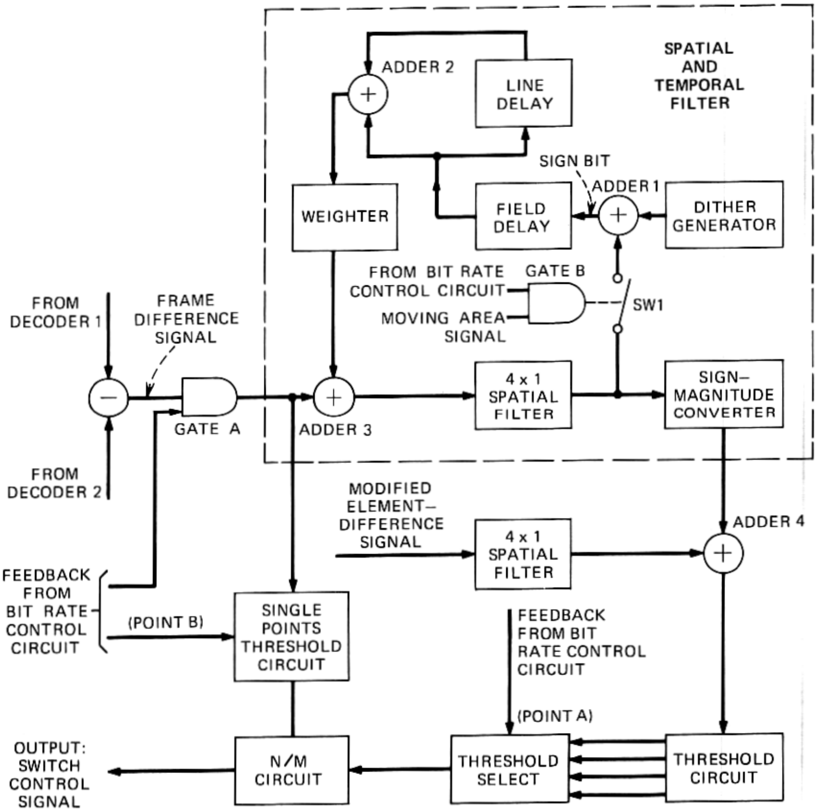


Fig. 10—Diagram of the moving area detector. The detector first filters the frame-difference signal spatially and temporally and then applies compensation for intraframe coding noise. The filtered signal is tested against one of several thresholds, and the resulting binary signal is blocked in using the N/M circuit. The output is used to select moving areas of the picture for transmission.

the coded-differential signal is added (adder 4) to the output of the sign-magnitude converter.[15] The purpose of this signal is to compensate for areas of the picture where more coding noise is likely to appear, namely at sharp edges where the outer decision levels of the quantizer are used.

Next, the output of adder 4 is fed to a circuit consisting of several thresholds. One of these thresholds is then chosen (depending on the bit-rate control strategy being used) as the input to an N/M circuit.[13] The function of the N/M circuit is to block in the moving area; that is, adjacent but noncontiguous points along a line are joined together to form one longer segment. In this way, the overall data rate is

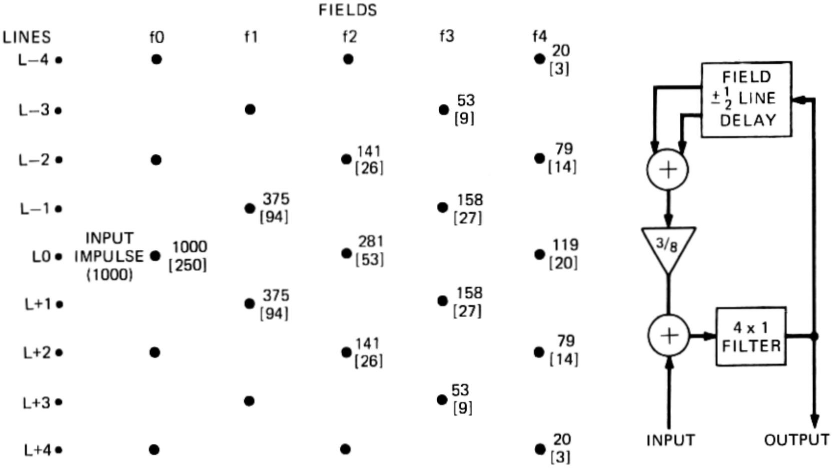


Fig. 11a—Impulse response of the spatio-temporal filter used in the moving area detector. The impulse response is a function of three dimensions, and the figure shows the response only in the vertical and temporal directions. The upper figures represent the area under the horizontal impulse response for each affected line in five fields. The lower (bracketed) figures represent the maximum value of each horizontal impulse response.

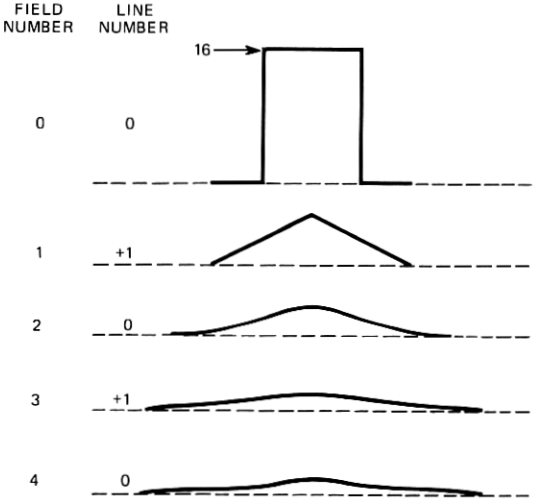


Fig. 11b—Horizontal impulse response of the spatio-temporal filter shown in Fig. 11a. The waveshape is given for line 0 and line 1 for each field of Fig. 11a.

reduced since each segment, however short, is allocated a 12-bit start-stop code, whereas the number of bits required to specify the amplitude of an element may be only 1 or 2.

By using the above-mentioned filtering techniques, large moving areas are easily detected; however, small isolated moving objects cause frame differences that, because of their short duration, are filtered out. Small moving objects of high contrast are detected by thresholding the unfiltered frame-difference signal with a large threshold value. The threshold signal is combined logically with the main signal path in the N/M circuit. The output of the N/M circuit is the final output of the moving area detector and controls the selection and transmission of new data (switches 1 and 2, Fig. 3).

A.3 Bit-rate control

The data-generation rate is matched to the transmission-bit rate by monitoring the level of fill of the transmission buffer and then applying controls to reduce the data-generation rate accordingly. These controls are applied to two parts of the system: the secondary element-differential encoder and the movement detector shown in Fig. 12.

A.3.1 Coder control

To reduce data in the encoder, a technique referred to as level-variable sampling is used.[17,18] The filtered energy in the error signal is important to the visibility of the quantizing error. Thus, close spacing of the inner levels insures that, where the input signal is fairly constant

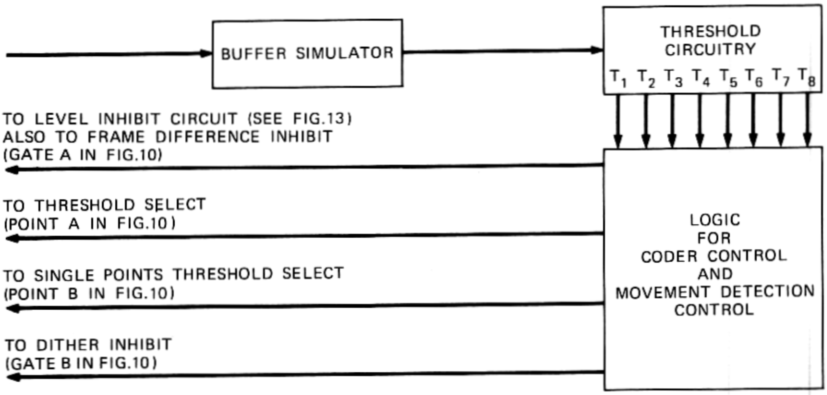


Fig. 12—Bit-rate control system. The system selects one or a combination of several bit-rate reduction techniques, depending on the level of the buffer simulator.

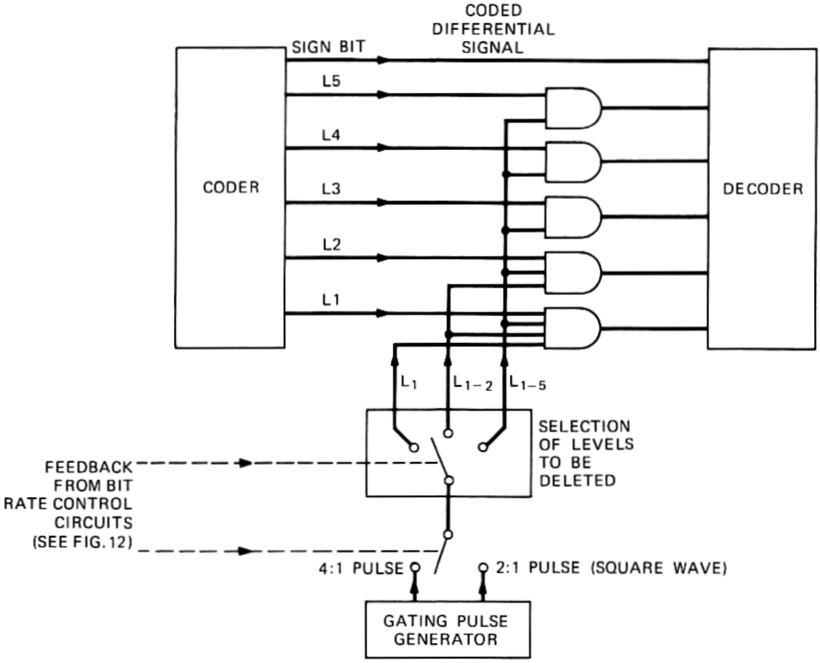


Fig. 13—Level-variable sampling. One or more of the coded differential level pairs is inhibited on every alternate element or, as an extreme measure, the signals are inhibited on three out of every four elements.

(low-detail area), the output signal will approximate the input very closely. However, such precision is not needed on every picture element. Consequently, the inner levels can be used less frequently than the outer levels.[17]

Figure 13 shows how level-variable sampling is performed. If, for example, we subsample just the inner pair of levels, L_1 is inhibited on every alternate element along the line. The effect of inhibiting a level is to change the quantizer scale, for that element, from an 11-level to a 9-level quantizer, as shown in Fig. 14. Two steps are taken to minimize the visibility of the resulting distortion: the subsampling pattern is synchronized to the horizontal rate; the pattern is staggered (by one element for 2:1 and two elements for 4:1 subsampling) so that subsampled elements are offset relative to the subsampled elements of the lines above and below (which are in the other field).

Control of the amount of data-rate reduction is achieved by switching between five different coder states. They are: (*i*) full sampling; (*ii*)

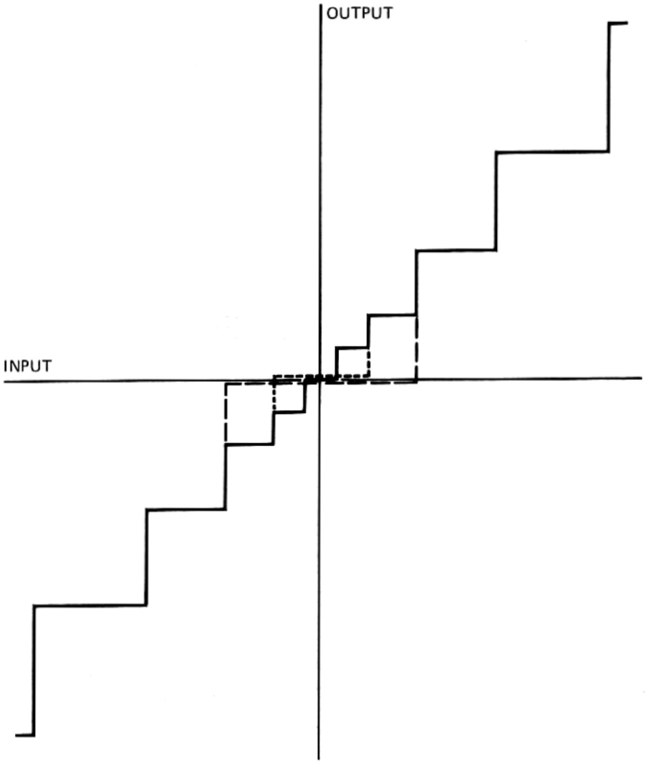


Fig. 14—Change in quantizer scale because of level variable sampling. Inhibiting the inner pair of coded differential levels (±1) changes the quantizer characteristics to that shown by the short dashed line. Inhibiting the two inner pairs of levels changes the characteristics to that shown by the long dashed line. On the elements in which no level inhibition takes place, the quantizer scale returns to normal (solid line).

subsampling on only the inner pair of levels (levels ±1); (*iii*) subsampling on the inner two pairs of levels (±1 and ±2); (*iv*) subsampling on all levels at a 2:1 rate; (*v*) subsampling on all levels at a 4:1 rate (1 element in 4 is sampled). A description of the variable word-length coding and the efficiencies achieved is given in Section A.3.3.

Subsampling introduces additional noise into the coding operation, particularly on the elements that are not sampled. This makes detection of the moving area more difficult especially since the signal from the primary coder is not subsampled. This problem is partially alleviated by setting the frame difference signal to zero on the unsampled elements by means of gate A in Fig. 10.

A.3.2 Movement detector control

As shown in Fig. 10, the movement-detector circuit generates a filtered frame-difference signal. Moving areas are detected by testing to see if the amplitude of this signal is above a certain threshold. By raising that threshold, less area will be detected as moving and less data will have to be transmitted. This is done, however, at the expense of reducing the quality of the picture in moving areas. Five different thresholds are used, as shown in Table II, so that the data rate can be reduced gradually.

Two other methods are used to reduce the amount of area detected by the movement detector: first, the feedback from the temporal filter is inhibited by gate B and switch 1 in Fig. 10, and second, the number of single points in the moving area is reduced.

The combined effect of the movement-detector controls is to gradually reduce the data rate and also to adapt the movement detector as the speed of movement increases so as to maximize its efficiency. Normally, when the subject is moving slowly, the amount of data being generated is small and the first mode of the movement-detector is used [i.e. (*i*) temporal filtering and (*ii*) low single-point threshold]. As the speed increases, the higher modes are used; the feedback from the temporal filter is inhibited, the filtered threshold is raised, and the single-point threshold is raised. Notice that the frame-difference signal resulting from faster movement is also larger.

A.3.3 Buffer simulator

A buffer simulator circuit was built to simulate operation at many different data rates. It is assumed that a variable-length code is used to transmit the coded differential signal. The lengths assigned to each classifier output are given in Table Va. Notice that the code changes depending on the particular coding mode that is being used. For example, for the mode where levels ±1 are deleted on alternate samples, the fully coded samples use code D with a maximum code-word length of 4 bits, whereas the alternate samples use code B with a maximum code-word length of 5.

Because the quantizer level usage changes from picture to picture, the codes of Tables Va and Vb will not always be optimum. In order to determine what could be gained by paying more attention to the code assignment (e.g., an adaptive strategy), we have calculated the efficiency of these codes for two different head-and-shoulders scenes. The entropy, bit rate, and efficiency for codes A, B, and C in one case and code D in another are given in Table VI. The results are given for four

Table Va — Four variable-word-length codes used in the CR/IR code

Level	Code Word Length			
	Code *A*	Code *B*	Code *C*	Code *D*
0	1	1	1	3
±1	3	—	—	3
±2	4	3	—	3
±3	5	4	3	4
±4	6	5	4	4
±5	6	5	4	4

Table Vb — The particular code used by each bit-rate control mode

Mode 1 Full Sampling	Modes 2, 3 Levels ± 1 Delete		Mode 4 Levels ± 1, ± 2 Delete		Modes 5, 6, 7 2:1 and 4:1
	Unconditional Samples	Conditional Samples	Unconditional Samples	Conditional Samples	
Code *A*	Code *D*	Code *B*	Code *D*	Code *C*	Code *D*

modes corresponding to (see Table II): (*i*) full sampling; (*ii*) deletion of levels ±1; (*iii*) deletion of levels ±1 and ±2; and (*iv*) 2:1 subsampling. The asterisk denotes the codes that are actually used in the implementation. Picture *X* has somewhat more detail than picture *Y*. It would have been slightly more efficient to use code *D* for mode 1 rather than code *A* for these particular pictures.

The entropies are rather high for an 11-level differential quantizer. The reason for this in mode 1 is that the moving area detector will update moving edges and highly detailed areas more frequently than low-detail areas, resulting in higher usage of the outer levels which, in turn, increases the first-order entropy. For the other modes, the tendency for the entropy of the unconditional picture elements to increase because of the deletion of levels on the alternate elements is almost balanced by the reduction in the amplitude of the element-to-element difference caused by camera integration.

In all cases except one, the efficiency of the variable-length code is greater than 90 percent. For the conditional elements in mode 4, the distribution is very peaked and the entropy is less than 1 bit per

Table VI — Entropy, bit rate, and efficiency for pictures X and Y

Mode	Scene	Entropy	Code A,B,C (as applicable)	Efficiency (%)	Code D	Efficiency (%)
1	X	2.946	3.246*	90.8	3.057	96.4
	Y	3.146	3.433*	91.6	3.297	95.4
2	X Unconditional	3.072	3.517	87.3	3.277*	93.7
	Y	3.344	3.847	86.9	3.396*	98.5
	X Conditional	2.080	2.147*	96.9	3.067	67.8
	Y	2.551	2.727*	93.5	3.210	79.5
4	X Unconditional	2.987	3.105	96.2	3.004*	96.2
	Y	2.972	3.304	90.0	3.253*	91.4
	X Conditional	0.822	1.301*	63.2	2.854	28.8
	Y	0.866	1.286*	67.3	3.000	28.9
5	X	3.146	3.752	83.8	3.383*	93.0
	Y	3.157	3.449	91.5	3.309*	95.4

* Code used in implementation.

element. To improve efficiency, it would be necessary to code elements in groups rather than singly. However, in this case the overall gain would be small.

The heart of the buffer simulator is an accumulator loop. For each transmitted sample the accumulator is incremented by an amount equal to the length of the corresponding code word. In addition, a count of 12 is added every time a new segment is transmitted to the receiver; this could be eight bits for a start-of-run address and four bits for an end-of-run code word. A count of 12 is also added to the accumulator at the start of each line to permit the decoder to synchronize at the start of line. No allocation is made for a start-of-frame code word (if a 50-bit code word were used, we would have to say that we are operating at 1.503 megabits per second rather than 1.500 megabits per second). The accumulator is decremented at a constant rate depending on the particular transmission rate that is being simulated. Thus, the output of the buffer simulator shows how full a buffer would be if it were actually used to transmit data to the receiver.

A circuit similar to this is used to monitor the data-generation rate. The accumulator is incremented with the same signal as the buffer simulator, but at the end of each line the contents are strobed into a commercial counter which enables us to integrate the data-rate count over any desired period.

A.3.4 Bit-rate control system

The bit-rate control system (Fig. 12) monitors the buffer simulator. The output range of the buffer is divided into eight regions, and a control mode is selected depending upon which region the buffer is in. Each mode uses a combination of the two previously described bit-rate reduction techniques (i.e., coder control and movement-detector control). The function of each mode is given in Table II.

In the lower modes, little or no level-dependent sampling occurs and the movement detector uses a low threshold. The movement detector completely covers the moving areas, but may also respond to a small amount of residual noise so that some stationary areas of the picture may also be detected. The result is that for limited subject activity a relatively large amount of data is generated and, correspondingly, the quality is little different from the primary encoder output.

As the buffer level increases, the intermediate modes (modes 2 to 4) are used. In these modes, level-variable sampling is used on the inner one or two pairs of levels; the moving-area detector operates on a higher threshold. As the buffer level increases further, the high modes (modes 5 to 7) are used. Subsampling is used on all classifier levels: at first in a 2:1 ratio and then finally (in mode 7) in a 4:1 ratio. The moving area detector coverage is reduced in two ways: first, the single point threshold is raised so that fewer single points are detected; second, in each consecutive mode the moving area detector threshold is raised. Normally, when the high modes are used it is because the subject is moving fast. Under these conditions the effects of these bit-rate reduction modes is somewhat masked because of the nature of human vision. Furthermore, since large frame-difference signals are generated, the moving area can still be accurately defined even though high moving-area-detector thresholds are used.

If the buffer level continues to rise, transmission of data is stopped. In this case, the receiver repeats the information from the previous frame (stored in its frame memory) until such time as the transmitter buffer level reduces sufficiently to allow new data to be transmitted.

REFERENCES

1. R. D. Kell, "Improvements Relating to Electronic Picture Transmission Systems," British Patent No. 341811, April 1929.
2. A. J. Seyler, "An Experimental Frame Difference Signal Generator for the Analysis of Television Signals," Proc. I.R.E. (Aust.), *24*, 1963, pp. 797–807.

3. F. W. Mounts, "A Video Encoding System Employing Conditional Picture-Element Replenishment," B.S.T.J., *48*, No. 7 (September 1969), pp. 2545–2554.
4. R. E. Graham, "Predictive Quantizing of Television Signals," IRE WESCON Convention Record, Pt. 4, 1958, pp. 142–157.
5. K. Teer, "Investigations into Redundancy and Possible Bandwidth Compression in Television Transmission," Philips. Res. Rep., *14* (1959), pp. 501–556 and *15* (1960), pp. 30–96.
6. J. C. Candy, M. A. Franke, B. G. Haskell, and F. W. Mounts, "Transmitting Television as Clusters of Frame-to-Frame Differences," B.S.T.J., *50*, No. 6 (July-August 1971), pp. 1889–1917.
7. J. O. Limb, "Buffering of Data Generated by the Coding of Moving Images," B.S.T.J., *51*, No. 1 (January 1972), pp. 239–259.
8. B. G. Haskell, J. O. Limb, and R. F. W. Pease, "Frame-to-Frame Redundancy Reduction System Which Transmits an Intraframe Coded Signal," U. S. Patent No. 3,767,847, October 23, 1973.
9. H. Wendt, Interframe-Codierung fur Videosignale Internat. Elektron. Rundschau, *27*, No. 1 (January 1973), pp. 2–6.
10. F. Kanaya, Nippon Telegraph and Telephone, private discussions.
11. C. C. Cutler, "Differential Quantization of Communication Signals," U. S. Patent No. 2,605,361, July 29, 1952.
12. J. O. Limb and F. W. Mounts, "Digital Differential Quantizer for Television," B.S.T.J., *48*, No. 7 (September 1969), pp. 2583–2599.
13. R. F. W. Pease and J. O. Limb, "Exchange of Spatial and Temporal Resolution in Television Coding," B.S.T.J., *50*, No. 1 (January 1971), pp. 191–200.
14. D. J. Connor and J. O. Limb, "Properties of Frame Difference Signals Generated by Moving Images," to be published.
15. B. G. Haskell, D. J. Connor, and F. W. Mounts, "A Frame-to-Frame *Picturephone* Coder for Signals Containing Differential Quantizing Noise," B.S.T.J., *52*, No. 1 (January 1973), pp. 35–51.
16. R. T. Bobilin, "A Study of Prefilters and Sampling Rate for Intraframe *Picturephone* Coders," unpublished memorandum.
17. J. O. Limb, "Picture Coding: The Use of a Viewer Model in Source Encoding," B.S.T.J., *52*, No. 8 (October 1973), pp. 1271–1302.
18. J. O. Limb, "Adaptive Encoding of Picture Signals," M.I.T. Conference on Bandwidth Compression, April, 1969; Published in T. S. Huang and O. J. Tretiak, *Picture Bandwidth Compression*, New York: Gordon and Breach, 1972, pp. 341–382.
19. J. O. Limb, "A Picture Coding Algorithm for the Merli Scan," IEEE Transactions on Communications, *COM-21*, No. 4 (April 1973), pp. 300–305.
20. R. F. W. Pease, "Conditional Vertical Subsampling—A Technique to Assist in the Coding of Television Signals," B.S.T.J., *51*, No. 4 (April 1972), pp. 787–802.
21. E. F. O'Neill, "TASI: Time Assignment Speech Interpolation," Bell Laboratories Record, *37*, No. 3 (March 1959), pp. 83–87.
22. B. G. Haskell, "Buffer and Channel Sharing by Several Interframe *Picturephone* Coders," B.S.T.J., *51*, No. 1 (January 1972), pp. 261–289.
23. E. R. Kretzmer, "Reduced Bandwidth Transmission System," U. S. Patent No. 2,949,505, August 16, 1960.
24. K. Fukushima and H. Ando, "Television Band Compression by Multimode Interpolation," J. Inst. Elec. Commun. Eng. Jap., *47*, 1964, pp. 55–64.
25. P. A. Wintz, "Transform Picture Coding," Proc. IEEE, *60*, No. 7 (July 1972), pp. 809–820.
26. J. C. Candy and F. W. Mounts, "Redundancy Reduction System with Data Editing," U. S. Patent No. 3,571,807, March 23, 1971.
27. D. J. Connor, "Techniques for Reducing the Visibility of Transmission Errors in Digitally Encoded Video Signals," IEEE Trans. Commun., *COM-21*, No. 3 (June 1973).